



PATENT ABSTRACTS OF JAPAN

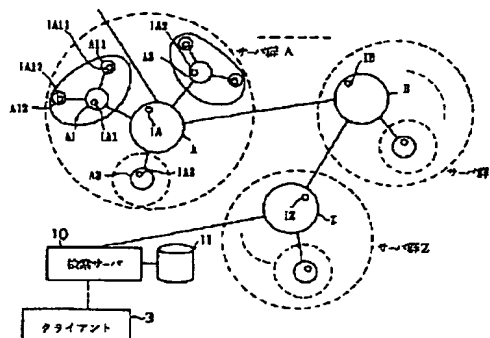
(11) Publication number: **10031683 A**(43) Date of publication of application: **03 . 02 . 98**(51) Int. Cl. **G06F 17/30**(21) Application number: **09001655**(22) Date of filing: **08 . 01 . 97**(30) Priority: **14 . 05 . 96 JP 08119213**(71) Applicant: **TOSHIBA CORP**(72) Inventor: **KUROSAWA HARUHIRO**

(54) **URL INFORMATION DATABASE CONSTRUCTING METHOD AND INTERNET RETRIEVAL SERVER UTILIZING THE METHOD** COPYRIGHT: (C)1998,JPO

(57) Abstract:

PROBLEM TO BE SOLVED: To construct a uniform resource locator(URL) information database by efficiently retrieving the URL information of a lot of world wide web(WWW) servers connected to a WWW.

SOLUTION: Concerning this URL information database constructing method, a lot of WWW servers connected to the WWW are divided into several groups A, B, C... and hierarchically structured in the respective groups and the WWW server A of much higher-order hierarchy is let preserve the URL information of all the WWW servers A1, A2... A11, A12... of low-order hierarchies connected to the present server. Thus, when a retrieval server 10 is to construct or to update a URL information database 11, access is performed to the most significant hierarchy WWW servers A, B, C... for each group divided into several groups so that the URL information of all the WWW servers can be collected.



JP-A 10-31683

[0034]

For example, a WWW server A11 collects and stores URL information on its server in a data area as data 11. Also, a WWW server A12 collects and stores URL information on its server in a data area as data 12. A WWW server A1 which belongs to a hierarchy layer upper than the servers A11 and A12 by one hierarchy layer collects and stores URL information on its own server (data 1), and the data (data 11) and the data (data 12) of the WWW servers A11 and A12 which belong to a hierarchy layer lower than the WWW server A1 by one hierarchy layer in a data area DATA. Furthermore, a WWW server A which belongs to the uppermost hierarchy layer collects and stores URL information on its own server (data 0), and the data (data 1), (data 11), (data 12); (data 2), (data 21), (data 22); (data 3), (data 31), (data 32)... contained in the respective data areas DATA of the WWW servers A1, A2, A3 which belongs to hierarchy layer lower than the WWW server A by one hierarchy layer in its data area DATA.

[0035]

This URL information collection is performed according to a procedure shown in a flowchart in Fig. 3. Actuation interval for the collection is not limited to a specific one such as each week, each month, each plural months, or each year. The interval is to be set so as to correspond to a data update frequency.

[0036]

Now, the WWW server A1 which belongs to an intermediate hierarchy layer will be explained. In a URL information collection function section IAl of the server A1, a self-information collection section Prog stores its own information in the data area DATA as data 1 (Step S1). Subsequently, whether or not URLs of a server group which belongs to a lower hierarchy layer have been registered in an additional information section URLR is determined (Step S2). When the determination is affirmative, http protocol issues URL of one server registered, for example, the URL of the server A11, according to a predetermined protocol (Step S3), the WWW server A1 receives the data (data 11) which has been stored in the data area DATA from the server A11 to store the data as data 11 together with self-data, data 1, in its data area DATA (Step S4).

[0037]

Thereafter, whether or not there is any URL which has been registered in the additional information section URLR but has not been accessed yet is determined (Step S2) again. When the determination is affirmative (here, the server A12 has not been accessed yet), the http protocol issues the URL of the server A12 (Step S3), the WWW server A1 receives the data (data 12) which has been stored in the data area DATA of the server A12 to store the data as data 12 in the self-data area DATA in an additional manner (Step S4). The above processing is repeated until access to all URLs which have been registered in the additional information section URLR but which have not been

accessed yet is completed.

[0038]

In this manner, regarding the WWW server A1 which belongs to the intermediate hierarchy layer, the data (data 1), (data 11), and (data 12) is collected and stored in the data area DATA according to the above URL information collection.

[0039]

Furthermore, in the case of an upper WWW server A, the servers A1, A2 and A3 are accessed to collect all the data (data 1), (data 11), (data 12); (data 2), (data 21), (data 22); (data 3),... in these data areas DATA and store a large volume of data shown in Fig. 2.

[0040]

A retrieval server 10 shown in Fig. 1 is accessible to these WWW servers on WWW, and it collects URL information of all the WWW servers which belong to lower hierarchies and which are connected to the retrieval server and URL information on the WWW servers A, B,...,Z by accessing respective URL information sections 1A, 1B,...,1Z of the WWW servers A, B,...,Z which are the uppermost hierarchy layer in the respective WWW server groups A, B,..., Z in order to obtain URL information for collecting in the URL information database 11.

[0041]

In this manner, by utilizing the retrieval server 10 which collects URL information on many WWW servers connected to WWW to store it in the URL information database 11, the Internet information retrieval system which performs information

retrieval configures a network shown in Fig. 4.

(19) 日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平 10-31683

(43) 公開日 平成10年(1998)2月3日

(51) Int. Cl. ⁶

G 0 6 F 17/30

識別記号

庁内整理番号

F I

G 0 6 F 15/40 3 7 0 Z
3 1 0 F

技術表示箇所

審査請求 未請求 請求項の数 5

O L

(全 1 1 頁)

(21) 出願番号 特願平9-1655

(22) 出願日 平成9年(1997)1月8日

(31) 優先権主張番号 特願平8-119213

(32) 優先日 平8(1996)5月14日

(33) 優先権主張国 日本 (J P)

(71) 出願人 000003078

株式会社東芝

神奈川県川崎市幸区堀川町72番地

(72) 発明者 黒澤 治弘

東京都港区芝浦一丁目1番1号 株式会社東

芝本社事務所内

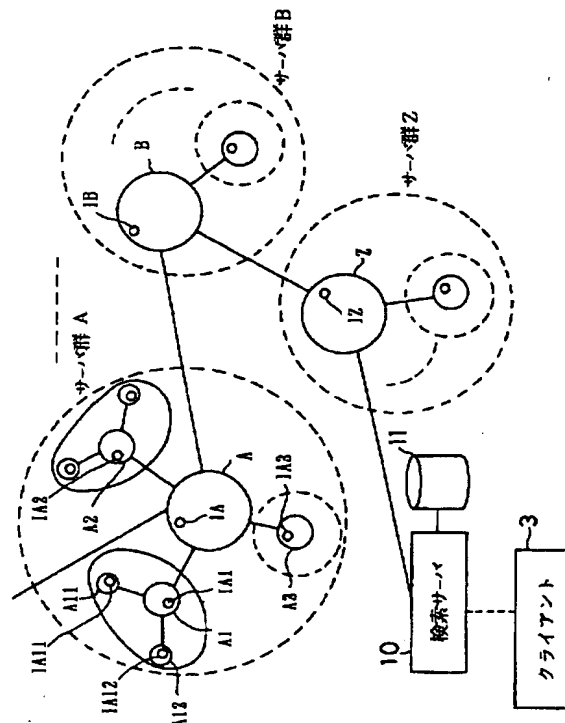
(74) 代理人 弁理士 三好 秀和 (外3名)

(54) 【発明の名称】 URL 情報データベース構築方法及びそれを利用したインターネット検索サーバ

(57) 【要約】

【課題】 WWWに接続されている多数のWWWサーバのURL情報を効率良く検索してURL情報データベースを構築することができるようにする。

【解決手段】 このURL情報データベース構築方法は、WWWに接続されている多数のWWWサーバをいくつかの群A, B, C, ...に分けて、それぞれの群内で階層構造化し、より上位階層のWWWサーバAに自サーバに接続されている下位階層のWWWサーバA1, A2, ...; A11, A12, ...すべてのURL情報を保持させることにより、検索サーバ10がURL情報データベース11を構築し、あるいはその更新を行う場合には、いくつかの群分けされた各群ごとの最上位階層のWWWサーバA, B, C, ...にアクセスすることによってすべてのWWWサーバのURL情報を収集できるようにする。



【特許請求の範囲】

【請求項1】 WWWに接続されている多数のWWWサーバを所定の規則にしたがって複数のWWWサーバ群の階層構造に分類し、

各階層におけるWWWサーバ各々に、自サーバに接続されている下位階層のWWWサーバについてそれらのURL情報を収集して保持させ、

前記複数のWWWサーバ群それぞれの最上位階層のWWWサーバに順次アクセスして、それらのWWWサーバが保持する自群に接続されている下位階層のWWWサーバのURL情報を収集してURL情報テーブルとして保存することを特徴とするURL情報データベース構築方法。

【請求項2】 前記各階層のWWWサーバが自サーバのURL情報と自サーバに属する1階層下位のサーバ群のURL情報とを記憶し、このURL情報に基づいてURLを発行し、受信したホームページ情報を記憶し、自サーバより1階層上位のサーバから前記URLを受信した時に自サーバに記憶している前記ホームページ情報を送信することを特徴とする請求項1に記載のURL情報データベース構築方法。

【請求項3】 クライアントに検索ホームページを発行する検索ホームページ発行部と、
前記クライアントが前記検索ホームページに書き込んだ検索条件を受信する検索条件受信部と、
前記検索条件受信部で受信した前記検索条件を解析し、検索条件要素を抽出する検索条件要素抽出部と、
請求項1又は2に記載のURL情報データベース構築方法を使用してURL情報を収集し、保持しているURL情報データベースと、
前記検索条件要素抽出部が抽出した検索条件要素について、前記URL情報データベースにアクセスして該当するURLを検索するURL検索部と、
前記URL検索部の検索結果に基づき、それぞれのURLアドレスを順次発行して該当するURLからの情報を収集するURL情報収集部と、
前記URL情報収集部が収集した前記URL情報を前記クライアントに送信するURL情報送信部とを備えて成る請求項1のURL情報データベース構築方法を利用したインターネット検索サーバ。

【請求項4】 前記URL検索部が検索した複数のURLについて、所定の優先条件に基づいてそれらのアドレス発行手順を決定し、その順序で前記URL情報収集部に渡すURL発行手順決定部を備えて成る請求項3に記載のインターネット検索サーバ。

【請求項5】 前記URL情報送信部が、前記URL情報収集部の収集した前記URL情報を所定の様式に編集する機能を備えて成る請求項4に記載の検索サーバ。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明はURL情報データベース構築方法及びそれを利用したインターネット検索サーバに関する。

【0002】

【従来の技術】 近年、インターネットの普及は急激に進んでおり、コンピュータの一般ユーザが手軽にWWW (World Wide Web) ブラウザを用いてアクセスしてWWW上に開いている種々のホームページを読み込み、必要な情報を得ることができる。

10 【0003】 図11は従来のインターネットと社内LANとの接続システムの構成を示しており、社内LAN1には社内サーバ2とこれに接続される、WWWブラウザ機能を有する多数のクライアント3があり、外部のインターネット4にはセキュリティも兼ねたゲートウェイ5を介して接続されている。WWWサーバ6は社内、社外にWWWサービスを提供するものである。

【0004】

20 【発明が解決しようとする課題】 インターネットの普及と共に世界中で多数のWWWサーバ6が接続され、それぞれのWWWサーバ6は独自のホームページを開いて、それにアクセスするためにはURL (Uniform Resource Locator) を指定する必要がある。

30 【0005】 ところが、図12に示すようにWWW上に開かれているホームページの数は無数に上り、あるクライアント3のユーザがWWWにアクセスして自分が真に必要なとしている情報を得ようとする場合には、必要とする情報を提供してくれるURLをWWW上で最初から特定するのは困難であり、例えば、最初に1つのWWWサーバ1のホームページH1にアクセスし、このホームページH1にハイパーリンク形式でつながっている関連するさらに下層のホームページH11, H12; H121, H122; H1221, ...などにアクセスしたり、あるいは他のWWWサーバ2, サーバ3, ...のホームページに移っていく方法をとらなければならない。

40 【0006】 そこで、このようにいわば手探り、やみくもな方法でWWWにアクセスするのでは、本当に必要とする情報を収集するには多くの時間がかかることになるため、市販されているいわゆるイエローページをあらかじめ参考にしてURLをピックアップしておき、そのURLにアクセスする方法をとるか、あるいはWWW上に提供されている検索サービスのホームページのURLにまずアクセスして、そこで自分が必要としている情報を提供してくれるURLを見つけ出し、その後、見つけ出したURLに再度アクセスするという方法をとることがある。

50 【0007】 しかしながら、この場合でもクライアントは自分で必要とする情報を提供してくれそうなホームページのURLを自分で検索し、そのURLをピックアップしてそれぞれにアクセスする必要がある、必要な情報

収集になお、時間と手間がかかる問題点があった。

【0008】このような問題点を解決するためにインターネット検索サーバを用意し、多数のWWWサーバのURL情報を収集してデータベース化して保持し、クライアントからの要望に応じてその検索条件を解析し、その検索条件要素を提供していると思われるURLを自動的に抽出し、それぞれのURLアドレスにアクセスしてホームページ情報を収集し、その結果を検索要求のあったクライアントに提供するインターネット検索サービスシステムが考えられている。

【0009】ところが、このようなインターネット検索サービスシステムでは、特に検索サーバに保持させるURL情報データベースのデータ内容をどのように収集するかが大きな課題になってくる。というのは、全世界には多数のWWWサーバが存在し、それらに開かれているURLは無数に上り、インターネットロボットを利用してそれらすべてにアクセスしてURL情報を収集してデータベース化するためには多大の時間と費用を要することになる。

【0010】本発明はこのような従来の技術的課題を解決するためになされたもので、世界規模で存在する多数のWWWサーバについていくつかのサーバ群に階層構造に群分けし、それぞれのサーバ群において上位階層のサーバが自サーバに接続されている下位階層のサーバそれぞれのURL情報を収集して保持するようにし、検索サーバからは複数の群分けされたそれぞれのサーバ群における最上位階層のWWWサーバにアクセスすることによって多数のURL情報を収集してURL情報データベースを構築することができるようにしたURL情報データベース構築方法を提供することを目的とする。

【0011】本発明はまた、このようにして検索サーバにおいて構築されているURL情報を利用して、クライアントから検索条件を入力させるための検索ホームページを発行し、そこにクライアントに検索条件を書き込ませ、その検索条件情報を受信した後は、検索条件から検索条件要素を自動的に抽出し、さらにURL情報データベースを活用にして該当するURLを抽出し、そのURLアドレスに自動的にアクセスして情報を自動収集し、検索結果をクライアントに送信する手順を取り、クライアントの要求に応じてきめ細かな検索情報サービスができるインターネット検索サーバを提供することを目的とする。

【0012】

【課題を解決するための手段】請求項1の発明のURL情報データベース構築方法は、WWWに接続されている多数のWWWサーバを所定の規則にしたがって複数のWWWサーバ群の階層構造に分類し、各階層におけるWWWサーバ各々に、自サーバに接続されている下位階層のWWWサーバについてそれらのURL情報を収集して保持させ、前記複数のWWWサーバ群それぞれの最上位階

層のWWWサーバに順次アクセスして、それらのWWWサーバが保持する自群に接続されている下位階層のWWWサーバのURL情報を収集してURL情報テーブルとして保存するようにしたものである。

【0013】この請求項1の発明のURL情報データベース構築方法では、WWWに接続されている多数のWWWサーバをいくつかの群に分けて、それぞれの群内で階層構造化し、より上位階層のWWWサーバが自サーバに接続されている下位階層の多数のWWWサーバのURL情報を保持させることにより、検索サーバにURL情報データベースを構築し、あるいはその更新を行う場合には、いくつかの群分けされた各群ごとの最上位階層のWWWサーバにアクセスすることによって多数のWWWサーバのURL情報が収集できるようになる。

【0014】したがって、WWWに接続されている多数のWWWサーバのURL情報を少ない数のWWWサーバにアクセスするだけで収集することができて、URL情報データベースの構築にかかる時間や費用を節約できることになる。

【0015】請求項2の発明は、請求項1のURL情報データベース構築方法において、前記各階層のWWWサーバが自サーバのURL情報と自サーバに属する1階層下位のサーバ群のURL情報とを記憶し、このURL情報に基づいてURLを発行し、受信したホームページ情報を記憶し、自サーバより1階層上位のサーバから前記URLを受信した時に自サーバに記憶している前記ホームページ情報を送信することを特徴とするものである。

【0016】この請求項2の発明のURL情報データベース構築方法では、各階層に属するWWWサーバ群の各々が自サーバに属する1階層下位のWWWサーバ群の各々にURLを発行してホームページ情報を受信し、自サーバのホームページ情報と共に記憶し、また上位階層のWWWサーバからURLを受信した時に当該ホームページ情報を送信する。したがって、階層構造に分類された多数WWWサーバのうち、最上位階層のサーバ群の各々にはそれらの属するすべての下位階層のWWWサーバ群のホームページ情報を収集して記憶していることになり、これらの最上位階層のサーバ群にアクセスしてそれらのホームページ情報を収集することによって大規模なURL情報のデータベースを容易に構築することができるようになる。

【0017】請求項3の発明のインターネット検索サーバは、クライアントに検索ホームページを発行する検索ホームページ発行部と、前記クライアントが前記検索ホームページに書き込んだ検索条件を受信する検索条件受信部と、前記検索条件受信部で受信した前記検索条件を解析し、検索条件要素を抽出する検索条件要素抽出部と、請求項1又は2のURL情報データベース構築方法を使用してURL情報を収集し、保持しているURL情報データベースと、前記検索条件要素抽出部が抽出した

検索条件要素について、前記URL情報データベースにアクセスして該当するURLを検索するURL検索部と、前記URL検索部の検索結果に基づき、それぞれのURLアドレスを順次発行して該当するURLからの情報を収集するURL情報収集部と、前記URL情報収集部が収集した前記URL情報を前記クライアントに送信するURL情報送信部とを備えたものである。

【0018】この請求項3の発明のインターネット検索サーバでは、クライアントから検索要求がインターネットを通じて送られてくると、まず検索ホームページ発行部がそのクライアントに検索ホームページを発行して、クライアントに検索条件を書き込ませる。

【0019】そこでクライアントが検索ホームページに書き込めば、その書き込まれた検索条件を検索条件受信部で受信し、検索条件要素抽出部によって検索条件を解析し、検索条件要素を抽出し、さらに検索条件要素抽出部が抽出した検索条件要素について、URL検索部がURL情報データベースにアクセスして関連する情報を提供しているURLを抽出する。そしてURL検索部が抽出したURLについて、URL情報収集部がそれぞれのURLアドレスを順次発行し、該当するURLからの情報を収集し、その収集結果をURL情報送信部によってクライアントに送信する。

【0020】こうして、クライアントから検索要求があれば、その検索条件を自身のホームページに書き込ませ、その検索条件要素を解析し、該当する情報を提供しているURLを抽出し、それらのURLに自動的にアクセスして情報を収集し、その結果をクライアントに提供することができ、クライアントのユーザとしては検索条件の指定だけで所望の情報を入手することができるようになり、インターネットを利用した情報収集がきわめて容易になる。

【0021】請求項4の発明は、請求項3のインターネット検索サーバにおいて、さらに、前記URL検索部が検索した複数のURLについて、所定の優先条件に基づいてそれらのアドレス発行手順を決定し、その順序で前記URL情報収集部に渡すURL発行手順決定部を備えたものである。

【0022】この請求項4の発明のインターネット検索サーバでは、URL検索部が検索した複数のURLについて、発行手順決定部が所定の優先順位決定条件に則ってアドレス発行手順を決定し、決定されたURLアドレスの発行手順に従ってURL情報収集部が順次該当するURLにアクセスして情報収集する。

【0023】したがって、検索されたすべてのURLについて無秩序にアクセスするよりも効率的な情報収集ができる。

【0024】請求項5の発明は、請求項4のインターネット検索サーバにおいて、前記URL情報送信部が、前記URL情報収集部の収集した前記URL情報を所定の

様式に編集する機能を備えたものである。

【0025】この請求項5の発明のインターネット検索サーバでは、URL情報収集部が収集した多数のURL情報をURL情報送信部で所定の様式に編集してクライアントに送信することができ、クライアントには検索条件に合致して必要な情報だけが分かりやすい形で配信できることになる。

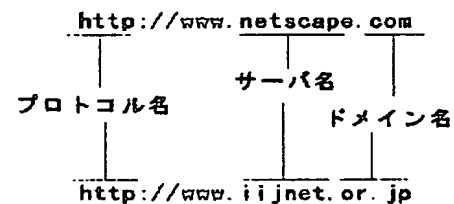
【0026】

【発明の実施の形態】以下、本発明の実施の形態を図に基づいて詳説する。図1は請求項1及び請求項2の発明のURL情報データベース構築方法の1つの実施の形態に用いるネットワークシステムを示しており、WWWに接続されているWWWサーバをいくつかのサーバ群A、B、C、…に群分けし、それぞれのサーバ群においてさらにいくつかの群A1、A2、…に分割し、分割された群内でさらにいくつかの群A11、A12、…；A21、A22、…；…に分割するという方法で階層構造に分割する。

【0027】このWWWサーバ群の分割手法は特に限定されないが、例えば、URLの記述構造を見ると、次の数1式のようにになっている。

【0028】

【数1】



このURLは世界的に重複を避けるために各国の専任機構が管理している。ドメイン名はサーバの分類種別を表し、co.jpは日本の会社、govは政府機関、orgは団体、or.jpは日本の団体というように定められている。したがって、これらのドメイン名をキーにしてサーバ群を階層構造に分類することができる。またサーバ名（プロバイダ名）をキーにして階層構造に分類することもできる。さらに、大分類にサーバ名を用い、その下位階層としてドメイン名を用いて分類する手法ならば物理的なネットワーク接続に基づく分類が可能となり、トラフィックが余分に発生せず、レスポンス的には優れたものとなる。

【0029】そして各WWWサーバは自サーバのURL情報を常に収集するURL情報収集機能部IA11、IA12、IA1、IA2、IA3、IA、IB、…を有するものとし、さらに上位階層のWWWサーバ、例えば、サーバ群Aについて考えると、子階層に当たるWWWサーバA1、A2、A3のURL情報収集機能部IA1、IA2、IA3それぞれは、孫階層に当たるWWWサーバA11、A12；A21、A22、…のURL情報収集機能部IA11、IA12；IA21、IA2

2, …それぞれが保持しているURL情報を収集して保持し、かつ自サーバのURL情報も保持している。

【0030】図2に示すように、このURL情報収集機能部IA11, IA12, IA1, IA2, IA3, IA, IB, …は、自己のURL情報を収集する自己情報収集部Prog、下位階層サーバのURLが格納してある付加情報部URLR、この付加情報部URLRの情報に基づいてhttpプロトコルが下位階層のWWWサーバにURLを発信し、収集した下位階層のURL情報とを記憶するデータエリアDATAから構成されている。

【0031】そしてURL情報収集機能部IA11, IA12, IA1, IA2, IA3, IA, IB, …それぞれは、自身の情報収集実行部hdtに登録されている周期、頻度で自己のURL情報収集と、自サーバの下位階層に属するWWWサーバ群のURL情報収集、更新処理を行う。すなわち、図1に示す場合、サーバ群Aにおいて最下位階層のWWWサーバA11, A12については、そのURL情報収集機能部IA11, IA12それぞれの付加情報部URLRの付加情報11、付加情報12は空であり、データエリアDATAには、自己情報収集部Progにより収集した自サーバのURL情報をdata11, data12として登録している。

【0032】これらのWWWサーバA11, A12の1階だけ上位階層のWWWサーバA1については、そのURL情報収集機能部IA1の付加情報部URLRの付加情報1として、1階だけ下位階層のWWWサーバA11, A12のURLが登録される。そして最上位階層のWWWサーバAのURL情報収集機能部IAの付加情報部URLRの付加情報0として、その1階だけ下位階層のWWWサーバA1, A2, A3のURLが登録される。

【0033】そして各階層のWWWサーバ各々は自己情報収集部Progによって自サーバのURL情報を収集し、またhttpプロトコルにより自サーバよりも1つの下位階層に属するWWWサーバ各々のURLを発信してそれらのデータエリアDATAに登録されている情報を収集し、自サーバのデータエリアDATAに登録する。

【0034】例えば、WWWサーバA11は自サーバのURL情報をデータエリアにdata11として収集、記憶している。またWWWサーバA12は自サーバのURL情報をデータエリアにdata12を収集、記憶している。そしてこれらのサーバA11, A12より1階だけ上位階層のWWWサーバA1は、そのデータエリアDATAに自サーバのURL情報data1と、1階だけ下位階層に属するWWWサーバA11, A12それぞれのデータエリアDATAのデータdata11, data12を収集して記憶する。さらに最上位階層のWWWサーバAは、そのデータエリアDATAに自サーバのURL情報data0と、1階だけ下位階層に属するW

WWサーバA1, A2, A3それぞれのデータエリアDATAのデータdata1, data11, data12; data2, data21, data22; data3, data31, data32…を収集して記憶しているのである。

【0035】このURL情報収集は、図3のフローチャートに示す手順で実行する。その起動間隔は週単位、月単位、複数月単位、あるいは年単位と特に限定されることはないが、データ更新頻度に応じた頻度に合わせることになる。

【0036】いま中間階層のWWWサーバA1について説明すると、このサーバA1のURL情報収集機能部IA1において、自己情報収集部Progが自己情報をデータエリアDATAにdata1として格納する(ステップS1)。続いて、付加情報部URLRに下位階層のサーバ群のURLが登録されているかどうか判断し(ステップS2)、登録されていればhttpプロトコルが所定のプロトコルにしたがって登録されている1つのサーバのURL、例えばサーバA11のURLを発行し(ステップS3)、そのサーバA11からデータエリアDATAに保存されているデータdata11を受信し、自身のデータエリアDATAにdata11として、自己データdata1と共に保存する(ステップS4)。

【0037】この後、再び付加情報部URLRに登録されているURLで未アクセスのものがどうか判断し(ステップS2)、未アクセスのURLがあれば(ここではサーバA12は未アクセスである)、httpプロトコルがそのサーバA12のURLを発行し(ステップS3)、そのサーバA12のデータエリアDATAに保存されているデータdata12を受信し、自身のデータエリアDATAにdata12として追加保存する(ステップS4)。以上の処理は付加情報部URLRに登録されているURLで未アクセスのものがなくなるまで繰り返す。

【0038】こうして中間階層のWWWサーバA1については、以上のURL情報収集によってそのデータエリアDATAにデータdata1, data11, data12を収集し、保存することになる。

【0039】さらに上位のWWWサーバAの場合には、サーバA1, A2, A3にアクセスしてそれらのデータエリアDATAのデータdata1, data11, data12; data2, data21, data22; data3, …をすべて収集して図2に示す大容量のデータを保存することになる。

【0040】図1に示す検索サーバ10は、WWW上でこれらのWWWサーバにアクセス可能であり、かつそのURL情報データベース11に収集するURL情報を得るために、各WWWサーバ群A, B, …, Zにおける最上位階層のWWWサーバA, B, …, ZのURL情報収

10

20

30

40

50

集機能部1A, 1B, ..., 1Zそれぞれにアクセスすることにより、それらのWWWサーバ自身のURL情報と共にそれらに接続されている下位階層のすべてのWWWサーバのURL情報を収集する。

【0041】このようにしてWWWに接続されている多数のWWWサーバのURL情報を収集してURL情報データベース11に保持している検索サーバ10を利用して、情報検索を行うインターネット情報検索システムは、図4に示すネットワークを構成する。

【0042】まず社内LAN1は社内サーバ2に多数のクライアント3を接続した形で構成されている。そして社内LAN1とインターネット4とはセキュリティの目的を兼ねたゲートウェイ5を介して接続されている。またこの社内LAN1にはインターネット4に接続するためのWWWサーバ6も用意されている。また上記のURL情報データベース11を有する検索サーバ10も接続されている。

【0043】請求項3～請求項5の発明の共通する1つの実施の形態のインターネット検索サーバ10は、図5に示す機能構成を有するコンピュータで構成される。検索サーバ10は検索条件要素となる多数のキーワードと、それらのキーワードに関連するURLとの対照テーブルで成るURL情報データベース11を備えている。一例を挙げれば、図8及び図9に示す内容である。図8はキーワードテーブル21であり、図9に示すURL情報テーブル22において使用されているキーワードをリストアップしたものである。したがって、このキーワードテーブル21にリストアップされていなければ検索不可とされることになる。図9に示すURL情報テーブル22は、多数のURLアドレスとそれぞれに関係があるキーワードとの対照テーブルである。例えば、

<http://www.tokyo.co.jp/sibaden>

というURLは、東京、渋谷にある量販店で通信販売も行っていて、取扱品目はパソコン関連商品全般ということになる。また価格リストも掲載していることを示している。

【0044】また、

<http://www.osaka.co.jp/sakaden>

というURLは、大阪、中之島にあるパソコン関連商品も取り扱っている家電製品の量販店で、通信販売も行っており、価格リストも掲載していることを示している。

【0045】そして検索サーバ10はこのURL情報データベース11のURL情報テーブル22を定期的に更新するが、そのためには上述した図1～図3に示したURL情報データベース構築方法を使用し、定期的にサーバ群A, B, ..., Zの最上位階層のWWWサーバA, B, ..., ZそれぞれにアクセスしてそれらのURL情報収集機能部1A, 1B, ..., 1ZからURL情報を収集する。この方法を利用することによって、各地に散在しているWWWサーバのすべてにアクセスしてURL情報

を収集するよりも格段に早くURL情報の収集が可能となる。

【0046】検索サーバ10はまた、クライアント3のユーザに検索要求手続を説明し、また検索条件を入力させるための検索ホームページを発行する検索ホームページ発行部12と、この検索ホームページ発行部12が発行する検索ホームページを通じてクライアント3から入力された検索条件を解析し、検索条件要素を抽出する検索条件要素抽出部13と、この検索条件要素抽出部13で抽出された検索条件要素に基づき、URL情報データベース11のキーワードテーブル21とURL情報テーブル22のキーワードリストを参照して、一致するキーワードがリストアップされているURLをピックアップするURL検索部14と、このURL検索部14がピックアップした複数のURLについて、あらかじめ設定されている優先条件に基づき、いずれのURLから発行するかの順序を決定するURL発行順序整理部15を備えている。

【0047】検索サーバ10はさらに、URL発行順序整理部15によって順序づけられた複数のURLについて、順次、URLを発行してWWWの該当するアドレスのホームページにアクセスするURL発行部16と、URL発行部16によってアクセスされたホームページそれぞれの情報を収集するURL情報収集部17と、収集したURL情報を編集してクライアント3に送信する検索結果編集部18と、そしてインターネットに接続して信号の送受信制御を行う送受信処理部19を備えている。

【0048】次に、上記構成のインターネット検索サーバの動作について説明する。図6に示すように、クライアント3から社内LAN1を通じて検索サーバ10のURLを指定してアクセスがあると、検索サーバ10の検索ホームページ発行部12は検索ホームページ（Home Page）を発行してクライアント3に送信する（ステップS11）。

【0049】クライアント3のユーザはこの検索ホームページの手続説明を読み、所定の手順に従って検索条件を入力する。いま、クライアント3から入力された検索条件は、図7のA1に示す内容であったとする。すなわち、「A地域内でB社のパソコンを購入したいので、各店の価格が知りたい。」

検索サーバ10の検索条件要素抽出部13では、クライアント3から検索条件を受信すると（ステップS12）、検索条件要素を抽出する（ステップS13）。ここでは、「A地域」、「B社パソコン」、「各店価格」という要素を抽出する（図7におけるA2）。

【0050】続いてURL検索部14がURL情報データベース11のキーワードテーブル21から検索条件要素に含まれるキーワードを抽出し、抽出したキーワードをリストアップしているURLをURL情報テーブル2

2からピックアップする(ステップS14)。ここでは、最適URLとしてURL1~URLXがピックアップされている(図7におけるA3)。

【0051】続いてURL発行順序整理部15においてURL検索部14がピックアップしたいいくつかのURL1~URLXについて、効率的な検索ができるように一定の優先条件に基づいてURLアドレスの発行手順を決定する(ステップS15)。この優先順序の決定には、クライアント3の検索条件要素に最も良く一致するキーワード数を有するURLから優先する方法が適当である。しかしながら、地域を優先する、技術分野を優先する、国名を優先するなど、あらかじめ優先条件を設定しておき、あるいはクライアント3からの検索条件入力時に、優先条件をユーザにチェックさせ、それに基づいて優先順位を設定する方法も採用することができる。A4はこうして決定された発行手順を示している。

【0052】この後、URL発行部16は決定された発行手順に従ってリストアップされているすべてのURLについて順次、URLアドレスを発行してそのホームページにアクセスし、そのホームページの情報を順次、URL情報収集部17に蓄積していく(ステップS16及び図7のA5)。

【0053】こうして所定のアドレス発行手順に従い、最後のURL3までの情報収集が完了すると、検索結果編集部18が収集情報を所定のレポート様式に編集してクライアント3に送信する(ステップS17、S18及び図7のA6)。

【0054】いま、上記のインターネット検索サーバの動作をさらに具体的に説明すると、次のようになる。クライアント3が検索サーバから発行された検索ホームページに東京地域で、B社パソコンの各店の売り価格を知りたいという検索条件を書き込んだとする。

【0055】これを受信した検索サーバでは、検索条件要素検出部13で「東京地域」、「パソコン」、「B社」、「価格」といった検索条件要素を抽出し、これをURL検索部14が受け取ると、「東京地域」に含まれる地域として「東京」、「新宿」、「渋谷」がリストアップされているURLを候補とし、「B社パソコン」に関連するキーワードとして「パソコン」あるいは「パーソナルコンピュータ」、「計算機」、「電子計算機」がリストアップされているURLを候補とし、さらに「価格」もリストアップされているURLを候補として抽出することになる。そこで、図9のテーブル22からは、URLa、URLbだけがピックアップされたとする。

【0056】次には、URL発行順序整理部15でこれらのURLa、URLbのどちらから優先させるかを決定するのであるが、この例では一致するキーワードの数も等しく、地域的にも近いので特に優先順位を付けず、抽出順に情報収集することにし、まずURLaのアドレスを発行してそのホームページの情報を収集し、続いて

URLbのアドレスを発行して情報を収集することになる。

【0057】この後、検索結果収集部18は図10に示すようなレポート形式に編集してクライアント3に送信することになる。なお、ここで「イ店」はURLaのアドレスのホームページを開いている店の名前で、「ロ店」はURLbのアドレスの店の名前である。

【0058】こうしてこの実施の形態のインターネット検索サーバによれば、クライアントのユーザはインターネット検索サーバのURLにアクセスし、開かれた検索ホームページに検索条件を入力するだけで、後はインターネット検索サーバ側で検索条件に合致するURLをピックアップして情報収集し、その結果をレポートとして得ることができるようになり、検索に要する時間、手間を格段に俊約することができるようになる。

【0059】なお、ここでインターネット検索サーバからのレポートはクライアント側が回線を接続したまま待つて得る方式であっても良いが、また検索条件を送信した後は回線をいったん切断し、電子メールの形でクライアントのメールアドレスに送信しておいてもらう方式をとることもできる。

【0060】

【発明の効果】以上のように請求項1の発明のURL情報データベース構築方法では、WWWに接続されている多数のWWWサーバをいくつかの群に分けて、それぞれの群内で階層構造化し、より上位階層のWWWサーバが自サーバに接続されている下位階層のWWWサーバのURL情報を保持させ、検索サーバにURL情報データベースを構築し、あるいはその更新を行う場合に、いくつかの群分けされた各群ごとの最上位階層のWWWサーバにアクセスすることによって多数のWWWサーバのURL情報を収集できるようにしているので、WWWに接続されている多数のWWWサーバのURL情報を少ない数のWWWサーバにアクセスするだけで収集することができて、URL情報データベースの構築にかかる時間や費用を節約できることになる。

【0061】請求項2の発明のURL情報データベース構築方法では、各階層に属するWWWサーバ群の各々が自サーバに属する1階層下位のWWWサーバ群の各々にURLを発行してホームページ情報を受信し、自サーバのホームページ情報と共に記憶し、また上位階層のWWWサーバからURLを受信した時に当該ホームページ情報を送信するので、階層構造に分類された多数WWWサーバのうち、最上位階層のサーバ群の各々にはそれらの属するすべての下位階層のWWWサーバ群のホームページ情報を収集して記憶させておくことができ、これらの最上位階層のサーバ群にアクセスしてそれらのホームページ情報を収集することによって大規模なURL情報のデータベースを容易に構築することができる。

【0062】請求項3の発明のインターネット検索サー

バによれば、クライアントから検索要求があれば、その検索条件を自身のホームページに書き込ませ、その検索条件要素を解析し、請求項1又は2の発明のURL情報データベース構築方法を使用して構築されたURL情報データベースを参照して該当する情報を提供しているURLを抽出し、それらのURLに自動的にアクセスして情報を収集し、その結果をクライアントに提供するようにしているので、クライアントのユーザとしては検索条件の指定だけで所望の情報を入手することができ、インターネットを利用した情報収集がきわめて容易になる。

【0063】請求項4の発明のインターネット検索サーバによれば、クライアントの検索条件に基づいてURL検索部が抽出した複数のURLについて、所定の優先順位決定条件に則ってアドレス発行手順を決定し、決定されたURLアドレスの発行手順に従って順次該当するURLにアクセスして情報収集するようにしているので、クライアントの検索条件に基づいて抽出されたすべてのURLについて無秩序にアクセスするよりも効率的な情報収集ができる。

【0064】請求項5の発明によれば、URL情報収集部が収集した多数のURL情報をURL情報送信部で所定の様式に編集してクライアントに送信することができ、クライアントには検索条件に合致して必要な情報だけが分かりやすい形で配信できる。

【図面の簡単な説明】

【図1】本発明のURL情報データベース構築方法の1つの実施の形態に使用されるインターネットのネットワーク構成を示す説明図。

【図2】上記の実施の形態に使用されるURL情報収集部のデータ構造図。

【図3】上記の実施の形態におけるURL情報収集手順のフローチャート。

【図4】本発明のインターネット検索サーバの1つの実施の形態のシステム構成図。

【図5】上記の実施の形態におけるインターネット検索サーバの機能ブロック図。

【図6】上記の実施の形態による検索処理のフローチャート。

【図7】上記の実施の形態による検索処理のフローチャ

ート。

【図8】上記の実施の形態におけるURL情報データベースのキーワードテーブルの説明図。

【図9】上記の実施の形態におけるURL情報データベースのURL情報テーブルの説明図。

【図10】上記の実施の形態における検索結果のレポートの説明図。

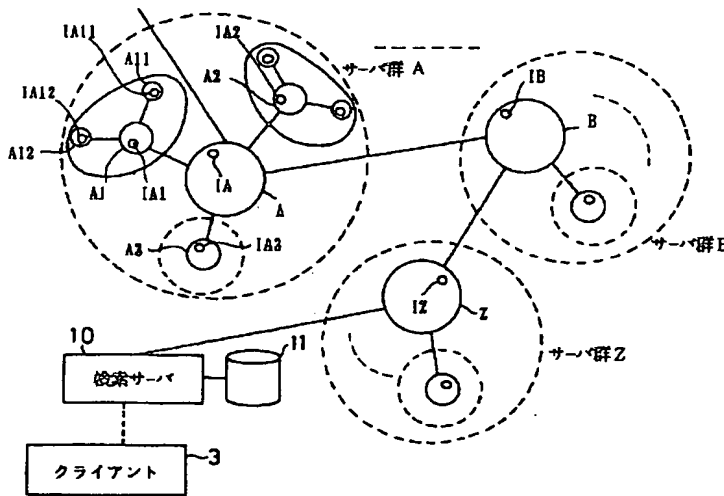
【図11】従来例のシステム構成図。

【図12】一般的なインターネットの構成図。

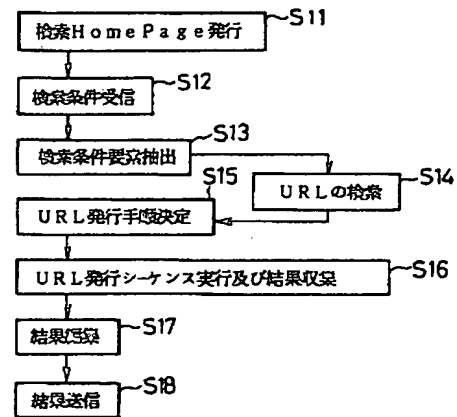
10 【符号の説明】

- 1 社内LAN
- 2 社内サーバ
- 3 クライアント
- 4 インターネット
- 5 ゲートウェイ
- 6 WWWサーバ
- 10 検索サーバ
- 11 URL情報データベース
- 12 検索ホームページ発行部
- 20 13 検索条件要素抽出部
- 14 URL検索部
- 15 URL発行順序整理部
- 16 URL発行部
- 17 URL情報収集部
- 18 検索結果編集部
- 19 送受信処理部
- 21 キーワードテーブル
- 22 URL情報テーブル
- A, B, C, ... WWWサーバ
- 30 A1, A2, A3, ... WWWサーバ
- A11, A12, A13, ... WWWサーバ
- IA, IB, IC, ... URL情報収集機能部
- IA1, IA2, IA3, ... URL情報収集機能部
- IA11, IA12, ... URL情報収集機能部
- Prog 自己情報収集部
- URLR 付加情報部
- http 情報収集実行部
- DATA データエリア

【図1】

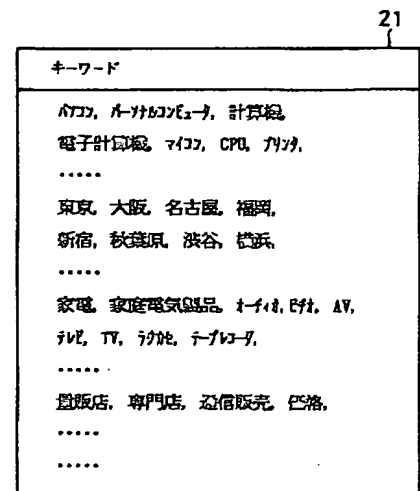
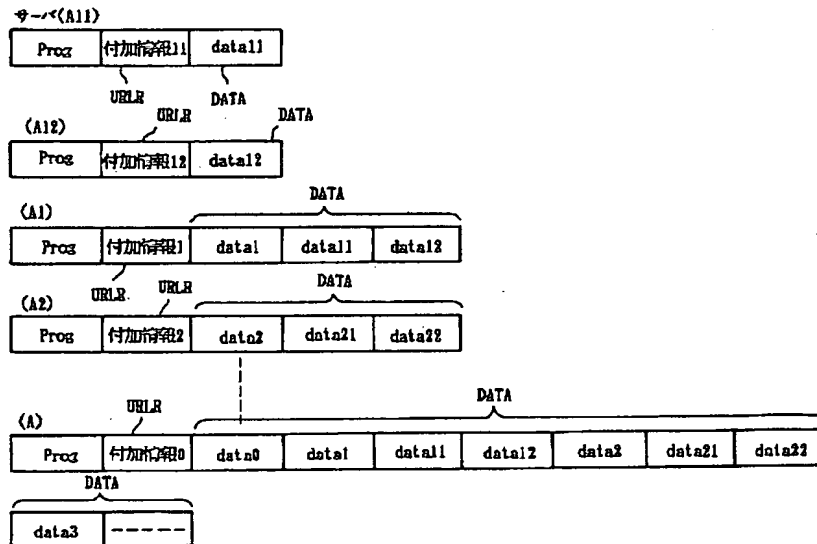


【図6】

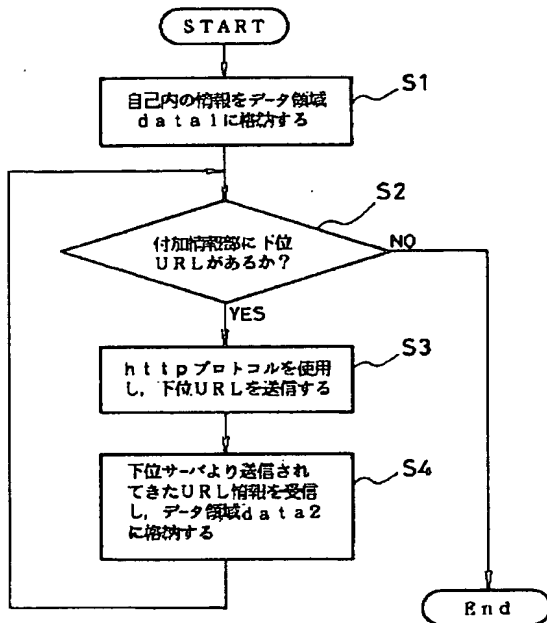


【図8】

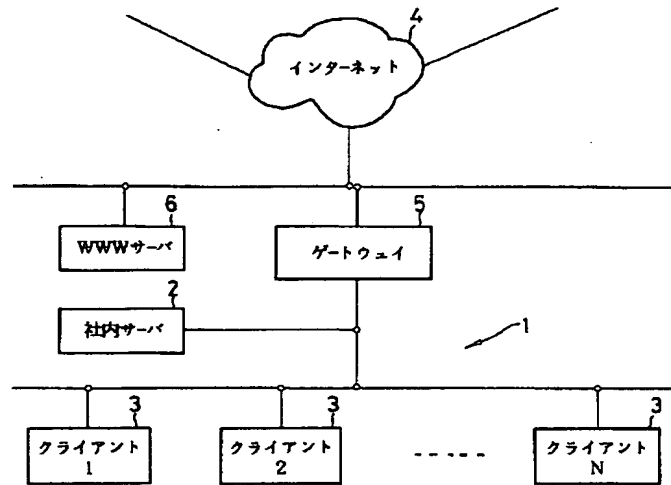
【図2】



【図3】



【図4】



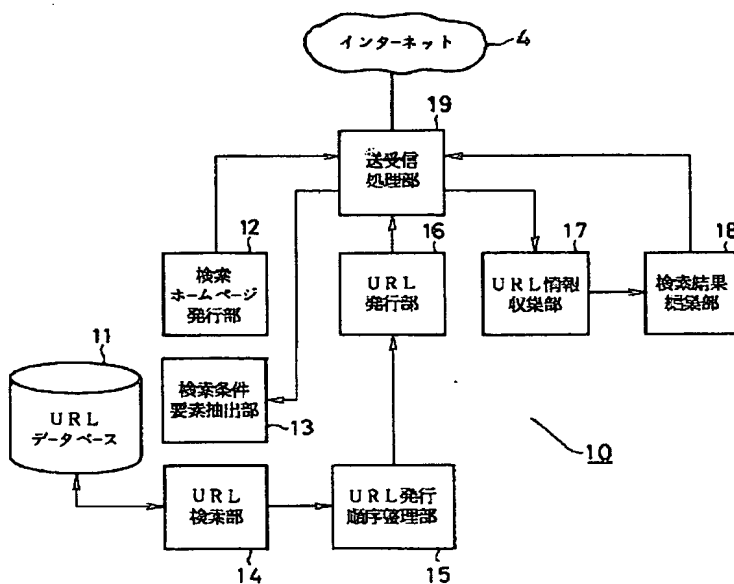
【図9】

URLリスト

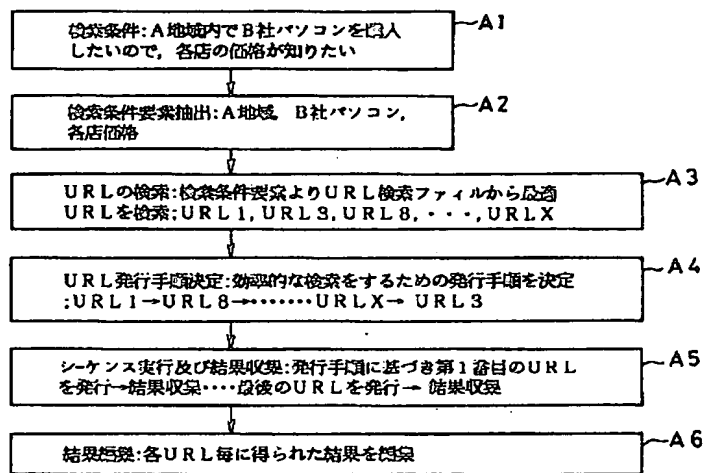
22

| | |
|-------|---|
| URLa | http://WWW.tokyo.co.jp/sibaden |
| キーワード | パソコン、パーソナルエディタ、計算機、パソコン、CPU、東京、渋谷、印刷点、通信販売、価格、... |
| URLb | http://WWW.kanto.or.jp/lohaden |
| キーワード | 家電、パソコン、CPU、マイク、東京、新宿、印刷店、価格、... |
| URLc | http://WWW.osaka.co.jp/sakaden |
| キーワード | パソコン、パソコン、マイク、テレビ、家電、大阪、中之島、印刷店、通信販売、価格、... |
| URLd | http://WWW.tokyo.co.jp/ysuscs |
| キーワード | 旅行、写真機、カメラ、カメラ、東京、新宿、印刷店、価格、... |
| ----- | |

【図5】



【図7】



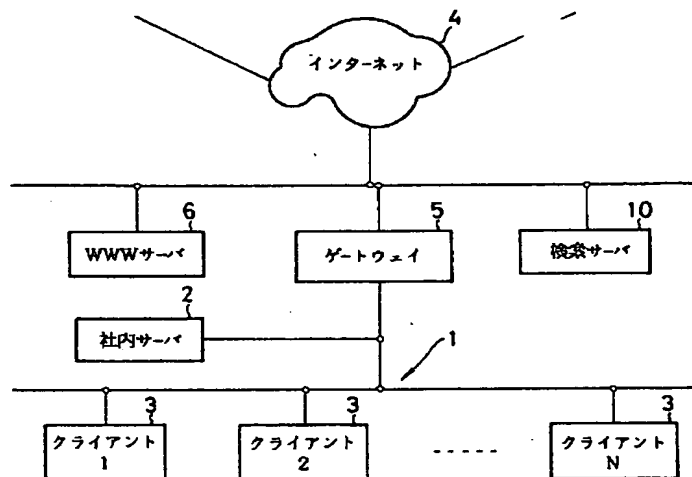
【図10】

ご依頼情報のサーチレポート

B社パソコンの価格リストは次の通りです。

| 型名 | イ店 | ロ店 |
|----|---------|---------|
| α | 253,000 | 264,000 |
| β | 324,000 | 314,000 |
| γ | 355,000 | 360,000 |
| δ | 425,800 | 415,800 |
| ε | 518,000 | 500,000 |

【図11】



【図12】

